A Study on SDN-based Pragmatic Service Management for Network Applications

著者	ギリエ バルハ ルイス アルベルト
journal or publication title	東北大学電通談話会記録
volume	89
number	1
page range	54-57
year	2020-08-31
URL	http://hdl.handle.net/10097/00128974

博士学位論文要約（令和２年３月）

SDN に基づくネットワークアプリケーションのための実用的サービス管理に関する研究

ギリエ バルハ ルイス アルベルト
指導教員：菅沼 拓夫，　学位論文指導教員：和泉 諭

A Study on SDN-based Pragmatic Service Management for Network Applications

Luis Alberto GUILLEN BARJA
Supervisor: Takuo SUGANUMA,　Research Advisor: Satoru IZUMI

Efficient networking is critical for the realization of Society 5.0. However, current best-effort network technologies have limited flexibility to assign resources. Additionally, Network Applications (NA), which have significantly influenced society by transforming users from simple content consumers to producers, also lead to the exponential growth of Internet traffic. Therefore, this study proposes pragmatic service management of NA, based on Software-defined Networking (SDN) as the core enabler. This dissertation presents various concepts on how Quality of Experience (QoE) and Service (QoS) on NA can be improved. For instance, Chapter 2 describes SAND/3, an SDN-based multi-level (i.e., network, transport, application, user) collaboration for resource assignment to improve the QoE of multimedia applications. Chapter 3 presents a hybrid load balancing mechanism to improve the QoS of NAs deployed over Content Delivery Networks using Distributed Storage Systems. Chapter 4 addresses a particular feature of QoS, service reliability, especially in cases of large-scale failure. Performance tests evaluated the proposed approach compared to standard methods using emulated and simulated environments, showing considerable improvements.

1. Introduction

The integration of the cyber-physical space will enable the realization of Society 5.0[1]. The main goal of Society 5.0 is to solve social problems by using cutting-edge technologies such as the Internet of Things (IoT), Big-data, Artificial Intelligence (AI), and robotics. This integration will also change the way people interact with each other and with their surroundings. Cyber-physical systems allow monitoring, control, and actuate in the real-world throughout services delivered via business applications, resulting in raw data that can be sent for further processing in a continuous cycle. However, having effective networking and communication is critical for this interaction.

From among those applications, Network Applications (NA)–i.e., programs that run on different end-systems and communicate over the Internet–have been the façade and the rationale of the Internet. Although in the early decades of the Internet, applications such as email, file transfer, or the World Wide Web (WWW) had a significant influence in society, in recent years Multimedia and Social Applications (MA) enabled the user paradigm-switch from simple consumers to producers, which influenced on how current and future Internet traffic will grow. It is foreseen that in a few years, about 80% will be video traffic not only from video services[2] but also other applications such as surveillance, gaming, or healthcare.

Services such as YouTube, Netflix, and Amazon are massively used worldwide; these MAs are deployed in distributed Content Delivery Networks (CDN), which are overlay networks comprised of servers that store the content in nodes geographically dispersed so that they can deliver the service closer to the end-user. However, these services still run on top of best-effort services, and therefore, having practical service management for MA is still needed.

In that regard, Software-defined Networking[3] (SDN) has gained popularity in recent years since it is changing the way networks are managed. SDN is an emerging network technology that allows decoupling the data plane (D-Plane), and the control/management plane (C/M-Plane). The Controller acts as the central entity that interacts with lower and higher-level applications, which allows having a centralized, programmable, and flexible control and overview of the entire network. Therefore, the primary hypothesis in this study is that SDN can help in the service management for NA.

The main goal of this dissertation is to realize pragmatic service management for NA, based on two major foundations: The collaboration of components, and using SDN as the enabler of that collaboration. In particular, we consider MAs due to its influence in future traffic. The target issues were addressed from two perspectives, the perceived user satisfaction of the service–Quality of

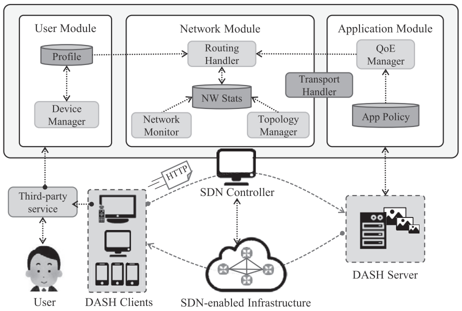

Figure 1: SAND/3 Architecture.

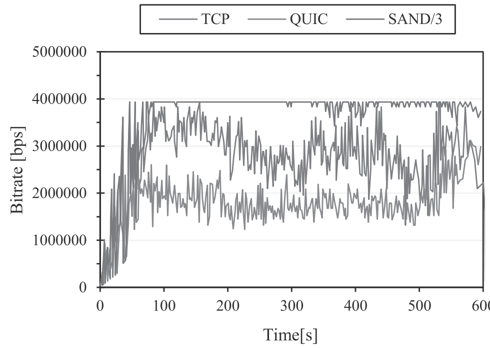

Figure 2: Average Media Throughput.

Experience (QoE)–and the performance–Quality of Service (QoS)–with particular focus on the network component. More specifically, in terms of QoE, Chapter 2 addresses multi-level collaboration for resource assignment. In the case of QoS, Chapter 3 addresses multi-level service load-balancing and its impact on other components. As part of the QoS, Chapter 4 addresses network reliability to improve service continuity; and finally, Chapter 5 draws some conclusions and presents some future directions of this study.

2. SDN-based Quality of Experience Management

QoE, the perceived user-satisfaction[4], has become an essential element to evaluate an application or service. In particular, for video streaming services, factors such as the video start delay, freezing events, quality shifts, or the media throughput (video quality) affect the QoE the most. Currently, Dynamic Adaptive Streaming over HTTP[5] (DASH) has become the de-facto standard for (stored and live) video content delivery. In DASH, media is cut into pieces (i.e., segments), which can be requested in its different representations–listed in a manifest file–over HTTP. Clients request a specific segment based on various parameters, such as the network status or the device specifications, in a continuous adaption cycle until all the segments of the video are played.

Since DASH does not define the implementation, there are various methods to perform the adaption, usually taking into account average parameters. However, since the adaption happens based on inaccurate information, and there is no control over the resource assignment, these are distributed unfairly among clients. Additionally, HTTP uses TCP as the transport protocol, which becomes a performance bottleneck, since it is not designed to reduce the delay.

In this chapter, the main issue addressed is the lack of collaboration of elements (i.e., network, transport, user) to handle the resource assignment. More concretely, the quality adaption does not consider the user requirements, so the QoE is not equally distributed among them.

Therefore, the primary goal of this chapter is to improve the QoE management for MAs using video streaming with DASH as the use-case. The overall architecture of the proposal, named SAND/3, is depicted in Figure 1. As observed, it consists of an SDN-based QoE management for DASH with multi-element collaboration for quality adaption using HTTP/3, whose major change is the usage of QUIC[6] as the underlying transport layer.

The overall process works as follows, based on a user profile, which contains the information gathered from the service provider and the devices, the proposed approach calculates a user category to send the segments via an adequate end-to-end route as indicated in the application policies.

To test the feasibility and effectiveness, we tested the approach on an emulated environment, and measure various QoE metrics (i.e., number of stalls, media throughput, and the number of quality shifts). For instance, Figure 2 shows the results obtained of the media throughput for a standard adaption bitrate algorithm over TCP, QUIC, and the proposed approach, where the latter obtained the steadiest average bitrate, which is desirable for better QoE. Therefore, based on the obtained results, this chapter showed how the proposed approach could pragmatically improve the QoE.

3. SDN-based Quality of Service Management

QoS, i.e., the characteristics of services to satisfy stated and implied user requirements[7], has been the core of the Service Layer Agreement (SLA). For instance, in networking, ensuring a minimum throughput is a QoS parameter. In the context of MAs, they are usually deployed in distributed CDNs and hosted in Data Center Networks (DCNs) provides a highly reliable infrastructure. However, when a failure occurs, moving data within and outside the DCN is costly, not only in terms of computation but also network bandwidth. Therefore, storage and network management are critical for QoS in MAs.

In that regard, Distributed Storage Systems[8] (DSS) may help solve storage reliability and performance. DSS link together unreliable nodes and combines network and storage capabilities to offer persistent and available service. However, the net-

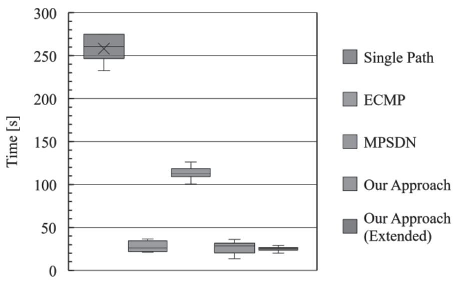

Figure 3: Completion Time per Approach.

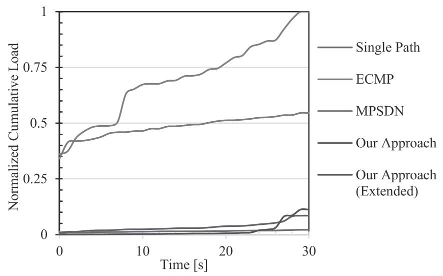

Figure 4: Cumulative Network Load.

work component is left to legacy techniques and technologies, which results in a performance bottleneck. Moreover, the routing has limited programmability leading to underuse of the available resources and creates an imbalance in the network and the service.

This chapter addressed three problems: the so-called last-mile bottleneck, which refers to the limited bandwidth at lower layers in the infrastructure, the limited use of multipath routing, and the load imbalance at the server and network level. Therefore, the main goal is to improve the QoS of MAs by enhancing the performance of CDNs with DSS.

The specific proposal consists of an SDN-based control method for DSS, with features such as on-demand inverse multiplexing, dynamic multipath routing, and a hybrid server-network load balancing to use the resources effectively. The overall proposed approach is comprised of modules that interact with service, network, and application. The most essential process takes part in the load balancing, which is done in a three-step sequence that discovers the available providers, assign them according to availability, and update them dynamically.

A performance test evaluated the proposed approach compared to standard methods using an emulated environment in various contexts. For instance, Figure 3 shows the results of an experiment that tested the completion time for a transmission, where the proposed approached showed a better performance. Moreover, Figure 4 shows how the proposed mechanism presents a much less cumulative load of the network, which will leave room for more effective use of the resources and,

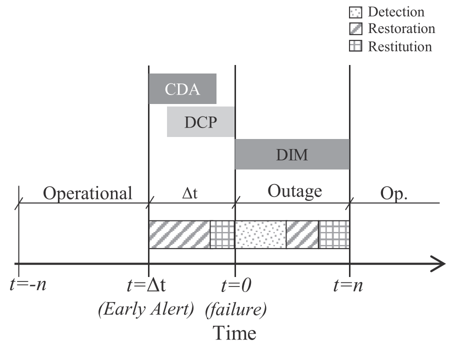

Figure 5: Proposed Timeline.

consequently, better QoS. Therefore, this chapter showed how the QoS could be improved in MAs by a centralized SDN-based control method that enhances the performance of CDN with DSS.

4. SDN-based Service Reliability Management

Service reliability of NAs will not be possible without network survivability[9)], which is the capability of a network to maintain the service continuity even in the presence of failures. Previous chapters showed how SDN offers various benefits to QoS and QoE; however, the SDN controller becomes a single-point-of-failure that affects service reliability. If a controller fails, even if the device is active, it is not programmable. Moreover, in the case of a large-scale failure due to an unexpected event (e.g., a disaster situation), a vast amount of devices is likely to fail–including controllers. Therefore, a prompt reaction is needed to maintain functional services.

However, most of the current solutions (either legacy or SDN-based) only consider single or few devices/link failures, but large-scale multi-controller failures are still not adequately tackled. Therefore, the main goal of this chapter is to improve the communication reliability immediately after a large-scale failure in SDN so that it is possible to avoid controller-device disruption and data-plane interference.

The proposed approach consists of a resilient controller communication in SDN by considering a predictive risk-avoidance, and a protection mechanism to mitigate the impact of the large-scale failure. Initially, we take advantage of the two possible ways to connect to a controller (i.e., in-band and out-of-band) and find a trade-off between restoration and protection mechanisms to protect the control line at the time of the early alert. The early alert might be of some seconds (in case of earthquakes) to a few days (in case of tornadoes) before the failure occurs. Therefore, the proposal considers all these factors and uses a three-stage protection mechanism, from the moment the early alert is triggered.

Figure 5 shows the proposed timeline, the first

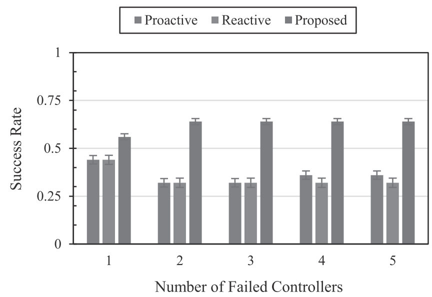

Figure 6: Completion Data Rate.

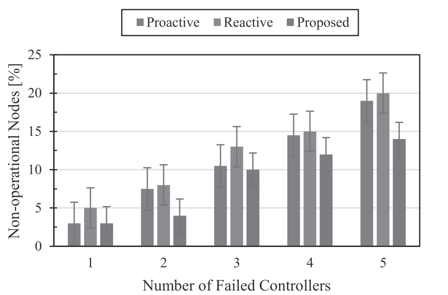

Figure 7: Percentage of Non-operational Nodes.

stage, called Controller Disconnection Avoidance (CDA), calculates a given number of alternative paths to control-lines around the disaster using in-band connections. Then the Risk Factor Index (RFI) assigns a value based on the distance to the failure event. The next step, called Data Communication Protection (DCP), assigns an RFI to all links to alert routing or higher-level mechanisms to be aware of the risk of each path at the data-plane. Finally, once the predicted failure starts to occur, the last step called Disaster Impact Monitoring (DIM) progressively updates the RFI every given interval.

We designed a model, implemented a simulator, and evaluated various parameters of the proposed approach. Figure 6 shows the results obtained at the data-plane, where, as observed, the proposed approach achieved a higher success rate of the transmissions. On the other hand, at the control-plane, Figure 7 shows how the proposed approach considerably reduced the percentage of non-operational devices, which consequently improves the overall service reliability.

To summarize, this chapter showed how resilient communication could be achieved in SDN for service continuity, especially when multiple controllers fail within a short period.

5. Conclusions

The main goal of this study is to realize pragmatic service management for network applications, which is based on two major foundations: The collaboration of components, and the usage of SDN as the enabler of such collaboration. The target issues are tackled from the QoE and QoS perspective. In the first case (QoE), we addressed the lack of collaboration of elements to handle the resource assignment and shown that the QoE of Applications can be improved by the proposed method, which uses multi-level collaboration orchestrated with SDN. In the second case (QoS), we addressed two features: service performance and reliability. By a hybrid load balancing control method, we could achieve higher performance, and by a three-step SDN controller communication, we showed how to improve reliability, even in cases where multiple controllers fail.

In the future, there are many directions in which this study can be extended. For instance, in Chapter 2, experimenting with multiple parallel segment delivery over HTTP/3 or using predictive segment request from content providers might be an exciting direction. In Chapter 3, using predictive path selection to optimize the route assignment, or using Network Function Virtualization (NFV) can further improve the applications. Finally, in Chapter 4, investigating the impact of multiple failure models in heterogeneous network environments would enhance the adaption scheme.

References

1) Japanese Cabinet, "What is Society 5.0?," `https://www8.cao.go.jp/cstp/english/society5_0/index.html` (2019-10-10).
2) CISCO, "The Zettabyte Era Trends and Analysis," White Paper. `https://files.ifi.uzh.ch/hilty/t/Literature_by_RQs/RQ\%20102/2015_Cisco_Zettabyte_Era.pdf` (2019-10-10).
3) Open Networking Foundation, "Software-Defined Networking Definition," `https://www.opennetworking.org/sdn-definition/` (2019-10-14).
4) Le Callet, P., Möller, S., and Perkis, A. (eds.) "Qualinet White Paper on Definitions of Quality of Experience," in *European Net. on Quality of Experience in Multimedia Sys. and Serv.*, Lausanne, Switzerland, 2013.
5) ISO/IEC 23009, "Information Technology–Dynamic Adaptive Streaming over HTTP (DASH)," *International Organization for Standardization (ISO)*, Geneva, Switzerland, 2012.
6) Langley, A. et al, "The QUIC transport protocol: Design and internet-scale deployment. *In Proc. ACM Special Interest Group on Data Communication*," LA–USA, pp. 183–196, 2017.
7) ITU-E.800, "Definitions of terms related to quality of service." `https://www.itu.int/rec/T-REC-E.800-200809-I` (2019-10-11).
8) Dimakis, A., et al. "Network coding for distributed storage systems," *IEEE Transactions on Information Theory*, Vol. 56, No. 9, pp. 4539–4551, 2010.
9) IETF RFC-3272, "Overview and Principles of Internet Traffic Engineering," `https://tools.ietf.org/html/rfc3272` (2019-11-11).